

An intelligent indian stock market forecasting system using LSTM deep learning

K. Kumar, Dattatray P. Gandhmal

School of Computer Science & Engineering, Vellore Institute of Technology, Vellore, India

Article Info

Article history:

Received Apr 20, 2020

Revised Jun 5, 2020

Accepted Jun 27, 2020

Keywords:

Classification
LSTM deep learning
Optimization
Stock data

ABSTRACT

Stock market data is considered to be one of the chaotic data in nature. Analyzing the stock market and predicting the stock market has been the area of interest among the researchers for a long time. In this paper, we have stepped forward and used a deep learning algorithm with classification to predict the behavior of the stock market. LSTM deep learning algorithm is used with an optimization algorithm to formulate the hyperparameters. To further improve the accuracy of prediction the stock data is first given to a classification algorithm to reduce the number of input parameters. In this research technical indicators are subjected to classification and deep LSTM algorithm which are both integrated to improve the accuracy of prediction. deep LSTM hyperparameters are trained using the optimization algorithm. In this paper Infosys and Zensar stocks data is collected from the Indian stock market data i.e. both national stock exchange (NSE) and Bombay stock exchange (BSE). The proposed approach is applied on Infosys and Zensar share values, the prediction accuracy obtained by employing this integrated approach of classification and LSTM has given a prominent value of MSE and RMSE as 1.034 and 1.002 respectively.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

K. Kumar

Department of Software Systems

Vellore Institute of Technology

Vellore Campus, Tiruvalam Rd, Katpadi, Vellore, Tamil Nadu 632014, India

Email: kkumar@vit.ac.in

1. INTRODUCTION

Securities exchange is the perplexing powerful framework and foreseeing the stock costs of things to come depends on the inner and the outside components. Despite the fact that a great deal of specialists inquired about on the expectation of the securities exchange costs, there is a huge variety in the conjecture contrasted and the first information. The way toward producing the high return to the securities exchange depends on the reality of bringing the compelling expectation with respect to the future appearance of budgetary resource costs. The financial exchange file as a theoretical arrangement of those stocks is a typical measure utilized for breaking down the presentation of the divisions, for example, the general securities exchange and a specific area [1, 2]. Then again, a market exchanging technique is guaranteed to be successful if and just if the exact forecast is guaranteed in that particular market. Forecast in the Stock market is by all accounts an incredible test in science both concerning the technique and dependent on the hypothesis of expectation [3, 4].

The forecast of the financial exchange sight and sound (graph) precisely is especially unrealistic utilizing the thoughts and strategies for the current scenario. The Efficient Market theory characterizes that all the present data are acquired in the stock costs and the nearness of new data cases to the capricious stock

costs. Irregular walk reports that the expectation of the stock costs isn't precisely anticipated by utilizing the authentic qualities representing the test that the Stock market forecast is a monotonous procedure [5, 6]. From the announcement given by the proficient market the past and current subtleties are acquired in the stock costs precisely, in this manner making us unmistakable that the adjustments in the cost are simply founded on the new data or "news", and doesn't rely upon the current data [7, 8]. The news in nature are available haphazardly and isn't known in the present, stock costs experience an irregular walk design and the present cost is the best wagered. When the endeavor to finish this idea turns out to be valid, the forecast of the financial exchange is futile.

Forecasting stock market: Till the present date, the strategies for forecast are sorted as four diverse types. The expectation techniques may be: time series method, regression method, expert-based method, and neural-network-based method [9, 10]. The forecast techniques require colossal measure of the past information and likewise, requires ordinary conveyances and utilizes the factual strategies that think about the framework qualities. This gains the expectation intense to ground and because of the cost the forecast appears to not material [11].

The most widely recognized techniques for the forecast in the budgetary markets incorporate the Time-arrangement investigation in which the expectations and choices are made profoundly dependent on the time-arrangement or the chronicled records of the stock costs [12, 13]. Most importantly, there are an enormous number of the variables that can't be acquired in the time arrangement however they have gigantic effects on the time arrangement. The proof in the financial exchange incorporates the changes in the cost of the stock that are chosen by the speculators [14]. The activities of the Investors' are especially nonsensical and they are basically reasonable and discerning dependent on the social association, social structure, aggregate convictions, and impression of this perplexing field. The huge issue behind the forecast is that on the off chance that it is conceivable to anticipate the modifications in the cost of the stock utilizing any of the extra, generally present, and significant information, alongside the time-arrangement information [15].

Impact of neural networks in stock market forecasting: One of the improvement instruments utilized for the forecast of the securities exchange dependent on the time series is the ANN that have the propensity to foresee the covered up or covered and obscure records. The most generally utilized territories of the securities exchange expectation utilizing ANN are zones of fund. In practically these applications, ANNs assume a significant job in learning the examples of the money related information. When the information is processed by ANN, the significant procedure utilized is the transformation of the information from its own numeric configuration to the numeric range that an ANN is fit for managing successfully. At this stage, the significance of changing the information symbolizes the learning procedure that target improving the generalizability of the scholarly outcomes. Another significant AI innovation is the Support Vector Regression and additionally, SVR calculation is utilized at anticipating the costs of the securities exchange [16, 17].

2. RELATED WORK

In this paper [18] author has evaluated hundreds of technical indicators and concluded that all the technical indicators are not always required. The author has applied Principal Component Analysis (PCA) and applied Hidden Markov Model to the selected features and obtained the required results.

In this paper [19] author has modelled the training algorithm on the Sensex Index by forecasting the volatility in the Indian Stock Market data. The author has applied various GARCH techniques and found that symmetric GARCH provided better results than any other techniques. He has also collected the data from 1996 to 2010 and applied the forecasting algorithm.

In this paper [20] author has proposed Neural Network utilizing the stock costs of Iran for a period of two years and the consequences of usage is appeared by legitimate charts. As a result, he demonstrated greater convenience of information mining in choice creation of financial exchange by choosing proper engineering to the Neural Network and getting ready information utilizing required and applicable procedures lastly preparing the system utilizing Back Propagation algorithm.

In this paper author [21] proposed model is a mix of information preprocessing techniques, hereditary calculations and Levenberg–Marquardt (LM) calculation for learning feed forward neural systems. In reality it advances neural system starting loads for tuning with LM calculation by utilizing hereditary calculation. The ability of the proposed technique was tried by applying it for anticipating some stock trade lists utilized. The outcomes show that the proposed approach can adapt to the changes of securities exchange esteems and furthermore yields great forecast accuracy.

In this paper [22] author proposed model Nonlinear Independent Component Analysis (NLICA), a novel element extraction system that accept the watched blends are non-direct mixes of inactive source signals, is utilized to discover free sources when watched information are blends of obscure sources, and

earlier learning of the blending instruments isn't accessible. In this paper, a stock value gauging model which first uses NLICA as preprocessing to concentrate highlights from anticipating factors is developed. The upside of the proposed strategy is that the data covered up in the first information can be found by highlight extraction. In this way, NLICA can give increasingly significant data to monetary forecasting.

In this study [23] author researched the probability of gauging an interim esteemed stock value record arrangement over short furthermore, long skylines utilizing multi-output support vector regression (MSVR). Besides, this examination proposes a firefly algorithm (FA)- based methodology, based on the built up MSVR, for deciding the parameters of MSVR (shortened as FA-MSVR). Three all inclusive exchanged wide market lists are utilized to look at the presentation of the proposed FA-MSVR technique with chose partners. The quantitative and complete appraisals are performed based on factual criteria, monetary criteria, and computational cost. Study [24] has focused on Stock market crash frame times and based on that various machine learning algorithms are applied to check the impact of crises on various factors in stock market such as bond, currencies and stock values etc. They have considered 39 countries stock data to analyse the effectiveness of the proposed approach.

From all the literature study it has been found that ensembled techniques are more effective compared to any single algorithm. Various challenges are still persisting in this study to suitably apply deep learning algorithm with some ensemble technique to enlarge the predicting value [25, 26]. Classification and then applying the machine learning algorithm shall be applied to further reduce the redundancy instead of directly passing the complete data the training algorithms.

3. PROPOSED WORK

Initially the Indian stock market data is obtained from the popular Indian Stock Exchanges such as National Stock Exchange (NSE) and Bombay Stock Exchange (BSE). The stock data of each company listed in the stock exchange consists of various parameters such as Open value, Close Value, High Value, Low Value, Volume, Adj. Close Value. The data is also available in various formats like Yearly, Monthly and Daily basis. The Various Technical indicators are applied on this stock data to understand the overall functioning of the stock market. Most stock market experts use these technical indicators to give any sort of prediction. Working Steps:

- Stock Data transformation to Technical Indicators
- Classification Algorithm on Technical Indicators
- Exploring LSTM algorithm on the classified data

As shown in Figure 1 the working of proposed system architecture, initially the stock data is given as input to system and transforms this stock data into technical indicators. All technical indicators are not always required further to remove redundant information wrapper based classification is applied and finally it is given to LSTM for training with help of optimization algorithm to generate the predicted values.

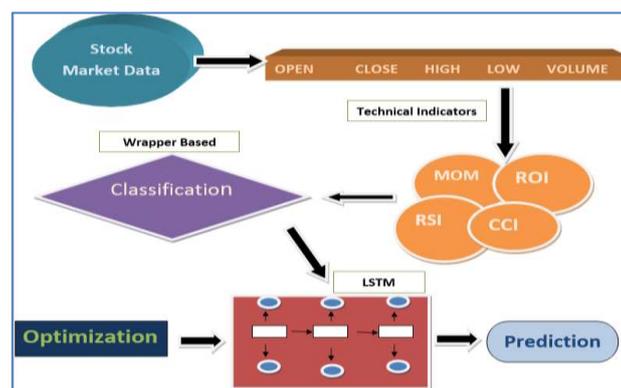


Figure 1. Proposed system architecture

3.1. Stock data transformation to technical indicators

Technical Indicators are the base for forecasting the Stock market data. Variety of Technical Indicators are available, the stock data is subjected to these technical indicators to transform into uniformity. Technical Indicators are the mathematical formulas applied on various parameters of the stock market to mostly used by the traders.

Often there is no defined rule to say that a particular technical indicator is best for the said company. The indicator which is best today may perform worst tomorrow, so to choose the right indicator is always a challenge to traders. Technical indicators such as WILLR, Rate of Change (ROCR), Momentum (MOM), Relative Strength Index (RSI) etc. are used to convert the overall input data into Technical Indicators format.

Rate of change (ROCR): It is used to calculate the rate of change with respect to the previous time intervals, which is denoted as,

$$R = \frac{P(r)}{P(r-a)} * 100$$

where, P denotes the price, $P(r)$ is the price at time r , and $P(r-a)$ denotes the price difference of study period time r , respectively.

Momentum (MOM): It is the measure used to compute the changes in price, which is expressed as,

$$M = P(r) - P(r-a)$$

3.2. Classification algorithm on technical indicators

These all huge variety of technical indicators are subjected to the classification algorithm to choose few technical indicators which can perform better on the selected company. Classification algorithms are very effective to test the credibility of the indicators and filter out the best out of it. The selected indicators are used as the important features for further applying the prediction algorithm. If the classifier is not applied this will lead to numerous redundant features resulting in average prediction. So, to improve the prediction rate classifier is applied and best features are only subjected to further process. Wrapper Approach is incorporated to identify the best Technical indicators suitable for prediction.

3.3. Exploring LSTM algorithm on the classified data

The selected features from the classification is given as an input to LSTM Deep learning algorithm. The hyper parameters of the LSTM are managed by optimization algorithms to correctly train the LSTM. RNN is not suitable for stock market prediction is because of its short memory remembering power. LSTM is known for remembering for long duration of time, this important feature of LSTM is very much applicable in stock market. Stock market values are also time series based since the previous days stock values are taken into consideration for forecasting the future prices. Hyperparameters of LSTM is also trained with the optimization techniques to further improve the price movement forecasting. In stock market the forecasting is based on short term and long term, same terms are coined using the LSTM to fit the requirements. Optimization algorithm also helps to solve the sparse gradient of Deep LSTM.

4. RESULTS AND DISCUSSION

4.1. Data set1

As explained in the above section the data of stock market is available as shown in Table 1 in multiple formats such as Daily, Monthly and Yearly. For the experiment we have taken the Infosys data from NSE and BSE. The data is collected from 1st Jan 2018 to 30th Nov 2019 shown in Table 1 is daily stock price having values of Open, High, Low, Close, Adj Close and Volume.

Table 1. Infosys stock market historical data (source: yahoo finance)

Date	Open	High	Low	Close	Adj Close	Volume
01-01-2018	518.85	522.25	515	516.775	479.5697	5431340
02-01-2018	518.625	521	511.5	514.85	477.7832	6112248
03-01-2018	514.25	515.8	509.3	510.65	473.8856	6846552
04-01-2018	510.5	510.5	504.8	507.7	471.148	8947614
05-01-2018	507.65	513.2	503	506	469.5703	11025976
08-01-2018	512.45	519.5	510.025	518.025	480.7296	5594690
09-01-2018	520	522.45	514.45	520.675	483.1889	9948236
10-01-2018	523.4	528	516.525	526.125	488.2465	9468634
11-01-2018	527.625	541.7	527.625	537.9	499.1737	16775298
12-01-2018	542.9	542.95	535.275	539.225	500.4033	18016528
15-01-2018	540	546.475	532.55	540.575	501.6561	10649250
16-01-2018	540.625	567.4	540.625	561.425	521.0051	22646966
17-01-2018	562.425	580.9	561.5	576.05	534.577	26154088
18-01-2018	572.5	582.45	566.05	576.35	534.8555	13204970
19-01-2018	572.5	573.95	568	571.725	530.5633	5988278

Figure 2 shows the graphical representation of Infosys stock price ranging from Jan 2018 to Nov 2019. The historical data of Infosys is forecasted as shown in Figure 3 using the proposed approach and the obtained results are highly significant. The MSE is 1.148 and RMSE is 1.163.

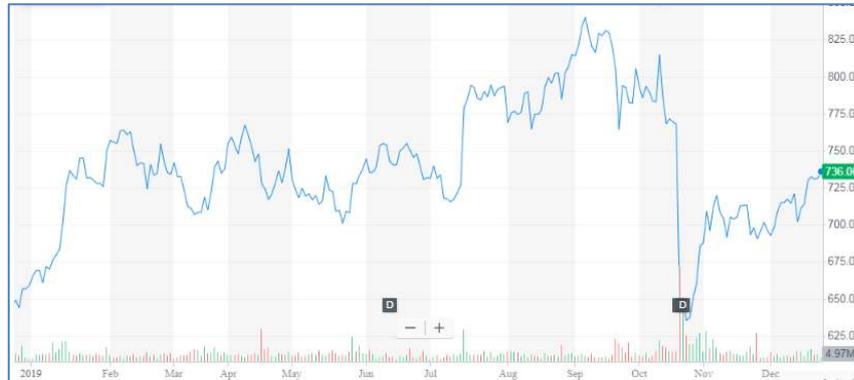


Figure 2. Infosys stock chart

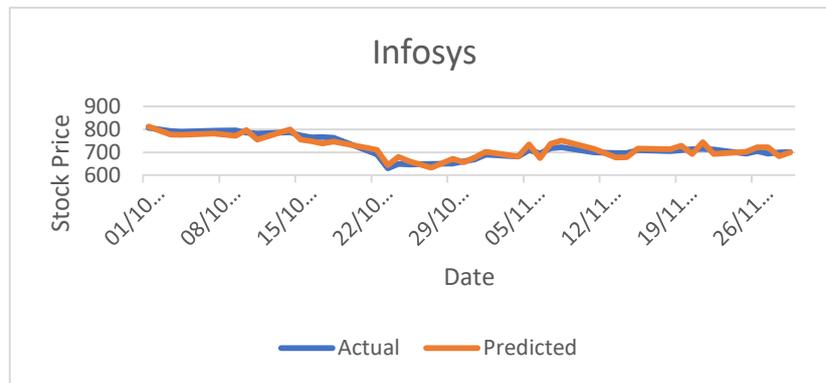


Figure 3. Infosys stock market forecasting

4.2. Data set2

For the experiment we have taken the Zensar data from NSE and BSE. The data is collected from 1st Jan 2018 to 30th Nov 2019 shown in Table 2 is daily stock price having values of Open, High, Low, Close, Adj Close and Volume.

Table 2. Zensar stock market historical data (source: yahoo finance)

Date	Open	High	Low	Close	Adj Close	Volume
01-01-2018	177.67	177.67	174.21	176.19	161.7906	3875
02-01-2018	176.19	176.19	170.9	172.21	158.1359	4380
03-01-2018	170.4	176	170.4	173.04	158.898	2540
04-01-2018	173.05	174.99	172	172.28	158.2001	1025
05-01-2018	173.99	174.4	170	173.57	159.3847	5175
08-01-2018	176	177.2	172	174.75	160.4683	4400
09-01-2018	179.94	179.99	176.4	178.96	164.3342	8610
10-01-2018	176.6	182	176.6	178.94	164.3159	9430
11-01-2018	180	184.8	180	183.89	168.8613	10890
12-01-2018	183.19	185.82	178.2	181.25	166.4371	4470
15-01-2018	183.93	183.93	179.4	180.01	165.2984	10285
16-01-2018	181.2	183.79	178.2	178.95	164.325	3185
17-01-2018	178.8	180.77	176.2	180.75	165.9779	17185
18-01-2018	181.38	186.8	177.58	179.68	164.9954	21355
19-01-2018	178.13	182.79	177.6	181.9	167.0339	18410
22-01-2018	180.2	189	180.2	187.51	172.1855	30900

Figure 4 shows the graphical representation of Zensar stock price ranging from Jan 2018 to Nov 2019. The historical data of Infosys is forecasted as shown in Figure 5 using the proposed approach and the obtained results are highly significant. The MSE is 1.034 and RMSE is 1.002.

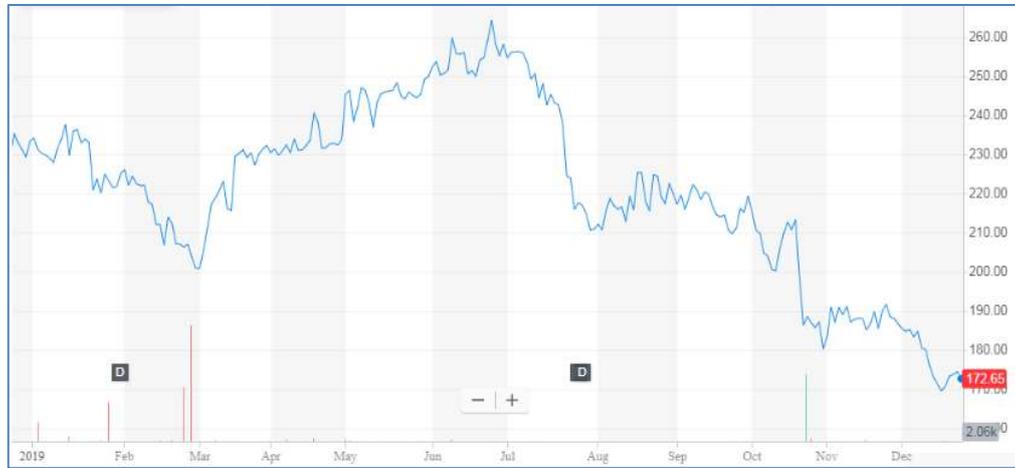


Figure 4. Zensar stock chart

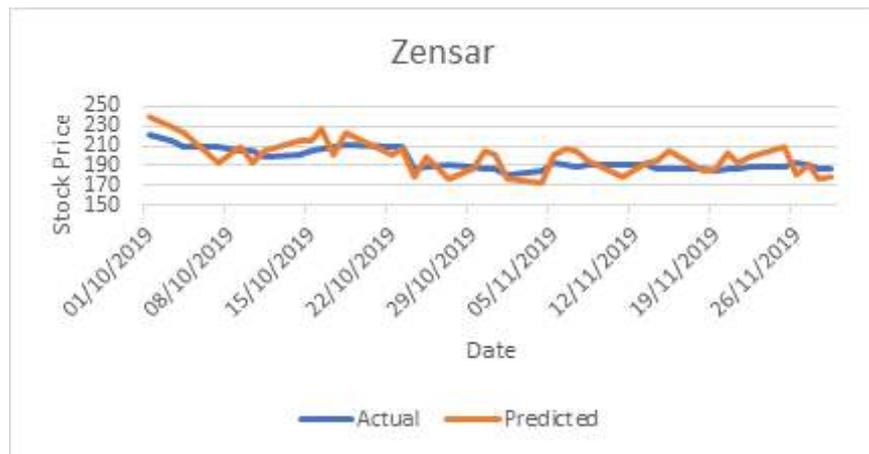


Figure 5. Zensar stock market forecasting

5. CONCLUSION AND FUTURE WORK

In this research a new model is proposed integrating the classification and LSTM algorithm for forecasting the Indian Stock Market data specifically NSE and BSE. The proposed approach is evaluated on company like Infosys and Zensar, it is observed that the obtained results are significant. The predictions provide will help the traders and stock market experts in making right decision whether to buy or sell a particular share. The results section also proves that the MSE and RMSE values obtained are optimum compared to any other approaches. The novelty proposed in this paper can be further extended by using various classifiers and training the model with a greater number of parameters.

REFERENCES

[1] Qiu, Mingyue, and Yu Song, "Predicting the direction of stock market index movement using an optimized artificial neural network model," *PloS one*, vol. 11, no. 5, 2016.
 [2] Xie, Xin-Kai, and Hong Wang, "Recurrent neural network for forecasting stock market trend," In *Computer Science, Technology and Application: Proceedings of the 2016 International Conference on Computer Science, Technology and Application (CSTA2016)*, pp. 397-402, 2017.

- [3] Rokade, Prakash P, "Business recommendation based on collaborative filtering and feature engineering–proposed approach," *International Journal of Electrical & Computer Engineering*, vol. 9, no. 4, pp. 2614-2619, 2019.
- [4] Zhou, Xingyu, Zhisong Pan, Guyu Hu, Siqi Tang, and Cheng Zhao, "Stock market prediction on high-frequency data using generative adversarial nets," *Mathematical Problems in Engineering*, 2018.
- [5] Zhang, Xi, Siyu Qu, Jieyun Huang, Binxing Fang, and Philip Yu. "Stock market prediction via multi-source multiple instance learning," *IEEE Access*, vol. 6, pp. 50720-50728, 2018.
- [6] Wahyono, Wahyono, Chasandra Puspitasari, Muhammad Dzulfikar Fauzi, Kasliono Kasliono, Wahyu Sri Mulyani, and Laksono Kurnianggoro, "An Optimal Stock Market Portfolio Proportion Model Using Genetic Algorithm," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 12, no. 2, pp. 171-180, 2018.
- [7] Saini, Deepak, and Akash Saxena, "Electric price forecast using interbreed approach of linear regression and SVM," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 2, no. 3, pp. 537-544, 2016.
- [8] Long, Wen, Zhichen Lu, and Lingxiao Cui, "Deep learning-based feature engineering for stock price movement prediction," *Knowledge-Based Systems*, vol. 164, pp. 163-173, 2019.
- [9] More, A. M., P. U. Rathod, R. H. Patil, D. R. Sarode, and B. E. Student, "Stock market prediction system using Hadoop," *International Journal of Engineering Science*, 16138, 2018.
- [10] Chong, Eunsuk, Chulwoo Han, and Frank C. Park, "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies," *Expert Systems with Applications*, vol. 83, pp. 187-205, 2017.
- [11] Khan, MD Hussain, and G. Pradeepini, "Machine learning based automotive forensic analysis for mobile applications using data mining," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 16, no. 2, pp. 350-354, 2015.
- [12] Wang, Feng, Yongquan Zhang, Qi Rao, Kangshun Li, and Hao Zhang, "Exploring mutual information-based sentimental analysis with kernel-based extreme learning machine for stock prediction," *soft computing* 21, no. 12, pp. 3193-3205, 2017.
- [13] Bharambe, Miss Minal P., and S. C. Dharmadhikari, "Stock Market Analysis Based on Artificial Neural Network with Big data," In *Proceedings of 8th Post Graduate Conference for Information Technology*, 2017.
- [14] Islam, Mohammad Rabiul, Imad Fakhri Al-Shaikhli, Rizal Bin Mohd Nor, and Vijayakumar Varadarajan, "Technical approach in text mining for stock market prediction: a systematic review," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 10, no. 2, pp. 770-777, 2018.
- [15] Sohangir, Sahar, Dingding Wang, Anna Pomeranets, and Taghi M. Khoshgoftaar, "Big Data: Deep Learning for financial sentiment analysis," *Journal of Big Data*, vol. 5, no. 1, 2018.
- [16] Chen, Weiling, Yan Zhang, Chai Kiat Yeo, Chiew Tong Lau, and Bu Sung Lee, "Stock market prediction using neural network through news on online social networks," In *2017 International Smart Cities Conference (ISC2)*, pp. 1-6. IEEE, 2017.
- [17] Ouahilal, Meryem, Mohammed El Mohajir, Mohamed Chahhou, and Badr Eddine El Mohajir, "A novel hybrid model based on Hodrick–Prescott filter and support vector regression algorithm for optimizing stock market price prediction," *Journal of Big Data*, vol. 4, no. 1, 2017.
- [18] Badge, Jyoti, "Forecasting of indian stock market by effective macro-economic factors and stochastic model," *Journal of Statistical and Econometric Methods*, vol. 1, no. 2, pp. 39-51, 2012.
- [19] Srinivasan, P., and P. Ibrahim, "Forecasting stock market volatility of BSE-30 index using GARCH models," *Asia Pacific Business Review*, vol. 6, no. 3, pp. 47-60, 2010.
- [20] Omid, Amir, Esmaeil Nourani, and Mehdi Jalili, "Forecasting stock prices using financial data mining and Neural Network," In *2011 3rd International Conference on Computer Research and Development*, vol. 3, pp. 242-246, 2011.
- [21] Asadi, Shahrokh, Esmaeil Hadavandi, Farhad Mehmanpazir, and Mohammad Masoud Nakhostin, "Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction," *Knowledge-Based Systems*, vol. 35, pp. 245-258, 2012.
- [22] Kao, Ling-Jing, Chih-Chou Chiu, Chi-Jie Lu, and Jung-Li Yang, "Integration of nonlinear independent component analysis and support vector regression for stock price forecasting," *Neurocomputing*, vol. 99, pp. 534-542, 2013.
- [23] Xiong, Tao, Yukun Bao, and Zhongyi Hu, "Multiple-output support vector regression with a firefly algorithm for interval-valued stock price index forecasting," *Knowledge-Based Systems*, vol. 55, pp. 87-100, 2014.
- [24] Chatzis, Sotirios P., Vassilis Siakoulis, Anastasios Petropoulos, Evangelos Stavroulakis, and Nikos Vlachogiannakis, "Forecasting stock market crisis events using deep and statistical machine learning techniques," *Expert Systems with Applications*, vol. 112, pp. 353-371, 2018.
- [25] Gandhmal, Dattatray P., and K. Kumar. "Systematic analysis and review of stock market prediction techniques," *Computer Science Review*, vol. 34, 2019.
- [26] Hiransha, M., E. Ab Gopalakrishnan, Vijay Krishna Menon, and K. P. Soman, "NSE stock market prediction using deep-learning models," *Procedia computer science*, vol. 132, pp. 1351-1362, 2018.

BIOGRAPHIES OF AUTHORS

Dr. K. Kumar is working in VIT-Vellore as Associate professor. He has completed his PhD in Software Engineering area. He has received his PG M.Tech.(CSE)from Pondicherry University and B.E.(CSE) from Madras University. He has 15 years of teaching/research experience and 3 years of industry experience in which he worked as a Software Engineer. His area of research includes AdvancedSoftware Engineering, Formalizing Requirement Patterns, Recommender system and design for engineering problems. He is also interested to work on Machine Learning and its usage in various systems. He has several publications on International conferences and Journals. He has expertise in accreditation and assessment.



Mr. Dattatray P. Gandhmal is pursuing his PhD in CSE from VIT, Vellore. He has completed his BE (CSE) in 2010 and ME (CSE) in 2012. He enjoys teaching and has 9+ years of Teaching Experience in Computer Science and Engineering discipline. He has Qualified GATE exam 5 times. His area of research includes Data Mining and Machine Learning. He has various publications on International conferences and Journals.